PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 3 NOVEMBER 15, 1917 Number 11

A NECESSARY AND SUFFICIENT CONDITION FOR THE EXISTENCE OF A STIELTJES INTEGRAL

By Gilbert Ames Bliss
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO
Communicated September 12, 1917

The purpose of this note is to suggest a proof of the following theorem:

If $u(x)$ is of limited variation and $f(x)$ bounded on the interval $a \leqq x \leqq b$, then a necessary and sufficient condition for the existence of the Stieltjes integral

$$\int_a^b f du \tag{1}$$

is that the total variation of $u(x)$ on the set D of discontinuities of $f(x)$ be zero. If $U(x)$ is the total variation of $u(x)$ on the interval ax, and α an interval with end points x_1, x_2, then by definition

$$U(\alpha) = U(x_2) - U(x_1)$$

and the total variation of $u(x)$ on D is defined to be the greatest lower bound of the sums

$$\sum_{k=1}^{\infty} U(\alpha_k)$$

for denumerable sequences of intervals $\{\alpha_k\}$ containing the points of D as interior points of the set which the intervals define.

Let the interval ab be subdivided by points x_i $(i = 0, 1, \ldots, n)$ with $x_0 = a$, $x_n = b$, $0 < x_i - x_{i-1} < \delta$ $(i = 1, 2, \ldots, n)$, and let S denote the sum

$$S = \sum_{i=1}^{n} f(X_i)\,[u(x_i) - u(x_{i-1})] = \sum_{i=1}^{n} f(X_i)\,\Delta_i u,$$

where the values X_i are arbitrarily selected in the intervals $x_{i-1}x_i$.

Then a necessary condition that the sums S have a limit as δ approaches zero is that

$$\lim_{\delta=0} \sum_{i=1}^{n} O_i |\Delta_i u| = 0, \tag{2}$$

where O_i is the oscillation of $f(x)$ on the interval $x_{i-1}x_i$. For it is clear that the values X_i can always be selected for two sums S, S' on the same partition so that the difference $S - S'$ is as near as is desired to the sum in (2). Hence if for every norm δ a partition can be found with

$$\sum_{i=1}^{n} O_i |\Delta_i u| > \eta > 0,$$

then for every δ two sums S, S' can be found with $|S - S'| > \eta$, and this would contradict the existence of the limit (1).

Let D_ϵ denote the closed set of points x where the oscillation $O(x)$ of f satisfies the inequality $O(x) \geqq \epsilon$, and let the intervals of S which contain points of D_ϵ as interior points be designated by the subscript j. Then it is also necessary that

$$\lim_{\delta=0} \sum_{j} |\Delta_j u| = 0. \tag{3}$$

For in every such interval $O_j \geqq \epsilon$, and if for every δ a sum as in (3) exists exceeding η, a sum as in (2) can be found exceeding $\epsilon\eta$.

Consider now a function $u(x)$ which is continuous as well as of limited variation, and suppose that the integral (1) exists. Then if δ is sufficiently small the j-intervals of S will be such that

$$\sum_{j} |\Delta_j u| < \eta.$$

Since $u(x)$ is now continuous the sum S can be altered by the introduction of new division points so that all the points of D are interior to the set of points defined by the sum of the j-intervals and the last inequality still true. Furthermore for the altered S it will follow that

$$\sum_{j} \Delta_j U < 2\eta$$

provided that δ is chosen so small that the total variation $U(b) - U(a)$, and $\sum_{i=1}^{n} |\Delta_i u|$, and hence also the corresponding quantities for every sum of partial intervals of ab, differ by less than η. Such a choice is always possible when $u(x)$ is continuous (Vallée Poussin, *Cours d'Analyse Infinitésimale*, vol. 1, 3d ed., p. 73). It follows then

that the total variation of u on the set D_ϵ must be zero, and hence also the total variation of u on the totality D of discontinuities of $f(x)$, since D is the sum of the sets $D_{1/n}$ $(n = 1, 2, \ldots)$.

To prove the sufficiency of the condition of the theorem for a function $u(x)$ which is continuous, consider the two monotonically increasing continuous functions $P(x)$, $N(x)$ satisfying the relations (Op. cit., p. 72)

$$u(x) - u(a) = P(x) - N(x),\ U(x) = P(x) + N(x). \qquad (4)$$

If the variation of $U(x)$ is zero on the set D then the same is clearly true of $P(x)$ and $N(x)$. Hence it is only necessary to prove the sufficiency for a monotonically increasing function $P(x)$. If D can be enclosed in the interior of a denumerably infinite system of intervals on which the sum of the variations of $P(x)$ is less than η the same is clearly true of its sub-set D_ϵ, and the latter is furthermore entirely interior to a finite number of the intervals since it is closed. (This follows from the Heine-Borel theorem. See, for example, Op. cit., p. 60, Lemme I. The proof is quite similar for a closed set or an interval.) The portion of ab remaining can be subdivided into intervals so small after D_ϵ is extracted that on each of them the oscillation of $f(x)$ is less than 2ϵ since on each such portion $O(x) < \epsilon$ (Op. cit., p. 254, §242). For every η and ϵ a partition of ab is defined in this way for which the inequality

$$\sum_{i=1}^{n} O_i \Delta_i P \leqq 2M\eta + 2\epsilon[P(b) - P(a)]$$

is true, where M is the upper bound of $|f|$ on ab. The first term on the right dominates the part of the sum for which the intervals contain points of D_ϵ, while the second does the same for the remaining terms. Hence the lower bound of the sum on the left is zero. But this is a sufficient condition that the integral

$$\int_a^b f\,dP$$

exists, as may be proved by precisely the method usually applied to the integral of Riemann (Op. cit., pp. 250, 255). It follows with the help of (4) that the integral (1) exists since the same is true when u is replaced by either P or N, and $u(x) - u(a) = P - N$.

The only case remaining to be considered is that of a function $u(x)$ of limited variation which has discontinuities. It is known that the totality of discontinuities of such a function is denumerable, and $u(x)$

is expressible as a sum $u(x) = v(x) + j(x)$, where $v(x)$ is of limited variation and continuous, and $j(x)$ is the 'function of jumps'

$$j(x) = \sum_{a<\xi\leq x} [u(\xi) - u(\xi - 0)] + \sum_{a\leq\xi<x} [u(\xi + 0) - u(\xi)]. \qquad (5)$$

The sums on the right are taken in each case for all the discontinuities ξ satisfying the inequality below the summation sign, and are absolutely convergent.

It can be shown first of all that if the integral (1) exists the discontinuities of f and u are necessarily distinct. For suppose there were a discontinuity ξ common to both with $|u(\xi + 0) - u(\xi - 0)| > \eta$, $O(\xi) > \eta$. On a sufficiently small interval $\xi - h \leq x \leq \xi + h$ the absolute value of $u(\xi + h) - u(\xi - h)$ and the oscillation of f would both exceed η. Hence for every norm δ there would exist two sums S and S' coinciding except on this interval and such that $|S - S'| > \eta^2$, which contradicts the existence of the limit (1). If $u(\xi - 0) = u(\xi + 0)$ at a discontinuity ξ, sums contradicting the existence of the limit can be similarly constructed by using intervals having ξ as end point.

In the second place the integral (1) with $j(x)$ in place of $u(x)$ always exists if the discontinuities of f and u are distinct, and it has the value

$$\int_a^b f\,dj = \sum_{\xi} f(\xi)\,[u(\xi + 0) - u(\xi - 0)], \qquad (6)$$

the sum being taken for all discontinuities ξ of $u(x)$ with the understanding that $u(a - 0) = u(a)$, $u(b + 0) = u(b)$. The series (5) for $x = b$ converge absolutely, and it is possible to select a set of discontinuities ξ_k $(k = 1, 2, \ldots, m)$ so that the sum of the absolute values of the terms of (5) not involving the ξ_k's is less than ϵ, and the corresponding sum from (6) less than ϵM. Consider now a sum S for the integral (6) with norm δ so small that no two values ξ_k can lie in the same interval. The value of S is the sum of the terms of $j(b)$ from (5) each multiplied by a factor of the form $f(X)$ with $|X - \xi| < \delta$. The terms not involving the ξ_k's have a sum less than ϵM in absolute value, while those involving the ξ_k's will differ from the corresponding terms of (6) by less than ϵM, say, if δ is sufficiently small since f is continuous at every ξ. Hence for a sufficiently small δ, the sum S differs from the sum on the right in (6) by less than $3\,\epsilon M$, which proves the statement at the beginning of this paragraph.

It follows then that a necessary and sufficient condition for the integral (1) to exist is that f and u have no discontinuity in common, and

that the integral shall exist with v in place of u. For since $u = v + j$ every sum S for u is the sum of two similar ones for v and j, and when f and u have no common discontinuity the last has a definite limit as δ approaches zero, as has been shown. This condition is, however, equivalent to that of the theorem. For the total variation of u on an interval is the sum of the total variations of v and j, and the fact that the set of points common to two denumerable sets of intervals is itself representable as such a denumerable set, makes it possible to show that the total variation of u on a set of points D is also the sum of the similar variations for v and j. Furthermore the total variation $J(x)$ of j is the value obtained from the series (5) by replacing each term by its absolute value, and on a set of points D containing no discontinuity of u it is zero. For if the ξ_k's are a set of points as described above, a denumerable set A of non-overlapping intervals can be selected approaching each ξ_k as a limit on both sides and containing all the points of ab except the ξ_k's. This set of intervals will also enclose D in its interior since D contains no ξ_k, and the sum of the total variations of j on A will surely be less than ϵ. It is easy to see, conversely, that when the variation of j is zero on D then the latter contains no discontinuity of u. Hence the existence of the integral with v in place of u and the conditions that f and u have no discontinuity in common imply that the total variation of u on D is zero, and conversely, which was to be proved.

TRANSFORMATIONS OF APPLICABLE CONJUGATE NETS OF CURVES ON SURFACES

By Luther Pfahler Eisenhart

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated by E. H. Moore, September 24, 1917

When the rectangular point coördinates x, y, z of a surface satisfy an equation of the form

$$\frac{\partial^2\theta}{\partial u \partial v} = a\frac{\partial\theta}{\partial u} + b\frac{\partial\theta}{\partial v}, \tag{1}$$

the curves $u = \text{const.}$, $v = \text{const.}$ form a conjugate system. We assume that the parametric system of curves is of this sort throughout this note, and we shall speak of it as a *net*. Equation (1) is the *point equation* of the net.

If N is such a net, the coördinates x', y', z' of a second net N' are given by quadratures of the form

$$\frac{\partial x'}{\partial u} = h\frac{\partial x}{\partial u}, \quad \frac{\partial x'}{\partial v} = l\frac{\partial x}{\partial v}, \tag{2}$$

provided that h and l are functions of u and v subject to the conditions

$$\frac{\partial h}{\partial v} = a(l-h), \quad \frac{\partial l}{\partial u} = b(h-l). \tag{3}$$

Moreover, each pair of solutions h, l of these equations leads by quadratures to a net N', such that the tangents at corresponding points M and M' to the curves of the net are parallel. All nets *parallel* to N are obtained in this way.

If θ is a solution of equation (1), and θ' is the corresponding function given by

$$\frac{\partial \theta'}{\partial u} = h\frac{\partial \theta}{\partial u}, \quad \frac{\partial \theta'}{\partial v} = l\frac{\partial \theta}{\partial v}, \tag{4}$$

then the functions x_1, y_1, z_1 defined by equations of the form

$$x_1 = x - \frac{\theta}{\theta'}x' \tag{5}$$

are the coördinates of a net N_1, so related to N that the lines joining corresponding points M and M_1 of these nets form a congruence whose developables meet the surfaces on which these nets lie in the curves of the nets. We say that the nets so related geometrically are in the relation of a *transformation* T. Parallel nets are in such relation. We have shown[1] that any transformation T of N into a non-parallel net N_1 is given by equations of the form (5). Hence any transformation T of N is determined by a parallel net and by a solution of the point equation of the net.

When two surfaces are applicable to one another, in the correspondence thus established there is a unique conjugate system of curves on one surface corresponding to a conjugate system on the other. We say that these nets are *applicable*. This paper is concerned with the transformations of applicable nets into applicable nets.

If we have two applicable nets N and $\overline{N}$ with the respective point coördinates x, y, z and $\bar{x}$, $\bar{y}$, $\bar{z}$, the analytical condition of their applicability is

$$\Sigma\left(\frac{\partial x}{\partial u}\right)^2 = \Sigma\left(\frac{\partial \bar{x}}{\partial u}\right)^2, \quad \Sigma\frac{\partial x}{\partial u}\frac{\partial x}{\partial v} = \Sigma\frac{\partial \bar{x}}{\partial u}\frac{\partial \bar{x}}{\partial v}, \quad \Sigma\left(\frac{\partial x}{\partial v}\right)^2 = \Sigma\left(\frac{\partial \bar{x}}{\partial v}\right)^2 \tag{6}$$

the sign Σ indicating the summation of terms in x, y and z. Since the coefficients a and b of equation (1) are functions of the left-hand

members of (6) and their derivatives, equation (1) is likewise the point equation of the net $\overline{N}$. In view of this fact a pair of functions h and l satisfying (3) leads to a net $\overline{N}'$ parallel to $\overline{N}$ as well as to N' parallel to $\overline{N}$. Moreover, the nets N' and $\overline{N}'$ are applicable. This result is due to Peterson.[2] The common point equation of the nets N' and $\overline{N}'$ admits the solution

$$\theta' = x'^2 + y'^2 + z'^2 - \bar{x}'^2 - \bar{y}'^2 - \bar{z}'^2. \tag{7}$$

This function and the corresponding function θ given by the quadrature (4) determine transforms N_1 and $\overline{N}_1$ of N and $\overline{N}$ respectively, and these new nets are applicable. Hence each net parallel to one of two applicable nets determines a parallel to the other, and by a further quadrature two applicable nets which are T transforms of the original nets. Moreover, the function θ' given by (7) is the only one leading to such a result. This result can be generalized at once to applicable nets in space of any order.

Nets are of three types with regard to applicability. Nets of the first type do not admit any applicable nets. Those of the second type admit one applicable net, whereas each net of the third type admits an infinity of applicable nets. We say that the latter are permanent in deformation, and for the sake of brevity call them *permanent nets*. Every net parallel to a permanent net is a permanent net, and each of the infinity of nets applicable to the one is parallel to one of the infinity applicable to the other by the method of Peterson. Suppose now that we have a permanent net N, two applicable nets $\overline{N}$ and $\overline{\overline{N}}$, and the respective parallel applicable nets N', $\overline{N}'$, $\overline{\overline{N}}'$. By the process of the preceding paragraph we obtain two transforms N_1 and N_2 of N, in general distinct, such that corresponding points of N, N_1 and N_2 lie on the same line, whose direction-parameters are the coördinates of N'. At the same time we obtain two transforms of $\overline{N}$ and two of $\overline{\overline{N}}$. As N admits an infinity of applicable nets, this process can be extended with the result that, in general, N and each of its deforms admit an infinity of transforms. We have raised the question whether in any case this infinity of transforms were coincident for each of the nets so that we obtain a permanent net N_1, whose infinity of applicable nets are the T transforms of the nets applicable to N. We refer to this question as *Problem A*.

Permanent nets belong to the general class of nets whose tangential coördinates satisfy an equation of Laplace with equal invariants. We have established[3] the existence of transformations T of nets of this kind into similar nets. When in particular the given N is a permanent net,

by the solution of a completely integrable system of partial differential equations of the first order a family of parallel nets of a particular type are obtained, each of which determines a T transform N_1, which also is a permanent net. All of these transformations are now shown to give a solution of Problem A.

In the transformations just referred to we did not consider permanent nets for which the curves in one family are represented on the Gauss sphere by one system of the imaginary generators. Drach[4] solved the problem of the deformation of nets of this kind. We show how in two ways these nets can be transformed into nets of the same kind as a solution of Problem A.

The third type of permanent nets are those whose two families of curves are represented on the sphere by its isotropic generators. These curves are the minimal lines on a minimal surface. There are no transformations of nets of this kind into similar nets furnishing a solution of Problem A.

[1] Eisenhart, *Trans. Amer. Math. Soc., New York*, **18**, 1917, (97-124).
[2] Peterson, *Ueber Curven und Flachen*, Moskau and Leipzig, 1868, (106).
[3] Eisenhart, *Rend. Circ. Mat., Palermo*, **39**, 1915, (153-176).
[4] Drach, *Ann. Fac. Sci. Toulouse*, (Ser. 2), **10**, 1908, (125-164).

ON BILINEAR AND N-LINEAR FUNCTIONALS

By Charles Albert Fischer

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY

Communicated by E. H. Moore, October 1, 1917

It has been proved by Riesz[1,2] that if a linear functional $A(f(s))$ is continuous with zeroth order, there is a unique regular function $\alpha(s)$ which satisfies the equation

$$A(f) = \int f(s)\, d\alpha(s),$$

and the variation of α is the least upper bound of the expression $|A(f)|/mf$, where mf is the maximum of $|f(s)|$. From this theorem Fréchet][3] has proved that if $U(f(s), g(t))$ is bilinear, that is linear in each argument, there is a function $u(s,t)$ which is regular in t and satisfies the equation

$$U(f,g) = \int\int f(s)\, g(t)\, d_2 u(s,t), \qquad (1)$$

and by modifying the definition of the variation of a function of two variables, he has proved that the variation of $u(s,t)$ is the least upper bound of $|U(f,g)|/mfmg$.

In the present note equation (1) is derived by a different method, and the function $u(s, t)$ is proved to be regular in both arguments when determined by this method, and unique if regular. The theorem can then be extended by mathematical induction to n-linear functionals.

The variation V_2u of $u(s, t)$ is defined by Fréchet as the least upper bound of the expression

$$\sum_{i,j} e_i e_j \, \Delta_{ij} u(s, t),$$

where e_i and e_j are taken equal to plus or minus 1 in such a way as to make the summation as large as possible, and

$$\Delta_{ij} u(s, t) = u(s_{i+1}, t_{j+1}) - u(s_i, t_{j+1}) - u(s_{j+1}, t_j) + u(s_i, t_j).$$

The double Stieltjes integrals considered here will all be of the form

$$\int_T f(s) \, g(t) \, d_2u(s, t) = \text{limit} \sum_{i,j} f(s_i') \, g(t_j') \, \Delta_{ij} u(s, t),$$

where the region T is defined by the inequalities

$$T: \; a \leqq s \leqq a'; \; b \leqq t \leqq b',$$

and s_i' and t_j' are in the intervals (s_i, s_{i+1}) and (t_j, t_{j+1}) respectively. Such an integral always exists if f and g are continuous and V_2u is finite, and it must satisfy the inequality

$$\left| \int_T f(s) \, g(t) \, d_2u(s, t) \right| \leqq mf \, mg \, V_2u.^3 \tag{2}$$

A function $u(s, t)$ will be called regular here if V_2u is finite, $u(a, t) = u(s, b) = 0$, and $u(s, t) = u(s+0, t) = u(s, t+0)$ excepting on the boundary of T. This makes some of the work simpler than to assume with Fréchet that $2u(s, t) = u(s-0, t) + u(s+0, t)$. If $u(s, t)$ is regular, its variation in one variable, when the other is constant, cannot be greater than V_2u. The double integral can then be expressed as an iterated integral by the equation[3]

$$\int_T f(s) \, g(t) \, d_2u(s, t) = \int_a^{a'} f(s) \, d_s \int_b^{b'} g(t) \, d_t u(s, t).$$

The functional

$$v(g(t); s) = \int_b^{b'} g(t) \, d_t u(s, t)$$

can be proved regular in s by proving that its variation cannot be greater than mgV_2u, and then proving that $v(g; s + \epsilon)$ approaches $v(g; s)$ when ϵ

approaches zero. The details of the proof of this will not be given here.

If two regular functions $u(s, t)$ and $u'(s, t)$ satisfy the equation

$$\int_T f(s)\, g(t)\, d_2 u(s, t) = \int_T f(s)\, g(t)\, d_2 u'(s, t) \tag{3}$$

identically in f and g, they will be proved to be identical. Since the similar theorem has been proved for single Stieltjes integrals, if u and u' are unequal at a point (s', t') there must be a continuous function $g(t)$ such that $v(g; s') \neq v'(g; s')$, where v' is the functional analogous to $v(g; s)$. Then since v and v' are regular in s there must be a continuous function $f(s)$ for which the two members of equation (3) are unequal. This proves that if a functional $U(f, g)$ satisfies the equation (1) where $u(s, t)$ is regular and independent of f and g, the function $u(s, t)$ must be unique.

As Riesz has proved,[2] the field of functions for which a linear functional is defined can be extended to any function which is the limit of a sequence of continuous functions which satisfy the inequalities

$$f_1 \geqq f_2 \geqq f_3 \geqq \ldots, \tag{4}$$

and any linear combination of such functions. Thus if $U(f, g)$ is bilinear, one or both of the functions f and g may be discontinuous if it is the limit of such a sequence. A function $f(s, s')$ will be defined by the equations

$$\begin{aligned} f(s, a) &= 0, \\ f(s, s') &= 1, \qquad (a \leqq s \leqq s';\ s' > a), \\ f(s, s') &= 0, \qquad (s' < s \leqq a'), \end{aligned}$$

and a function $g(t, t')$ by the analogous equations. If f is considered a function of s it is approached by a sequence such as (4), as Riesz has shown,[2] and $g(t, t')$ has the same property. The function $u(s, t)$ will then be defined by the equation

$$u(s', t') = U(f(s, s'), g(t, t')).$$

This function vanishes for $s' = a$, or $t' = b'$, by definition. Since U is bilinear the expression $U(f, g)/mfmg$ is bounded[3] and its least upper bound will be called M, a constant independent of f and g. The variation of u is defined as the upper bound of the expression

$$\sum_{i,j} e_i e_j \Delta_{ij} u(s', t') = U\left(\sum_i e_i (f(s, s'_{i+1}) - f(s, s'_i)) \sum_j e_j (g(t, t'_{j+1}) - g(t, t'_j))\right),$$

and since from definition

$$\sum e_i \left(f(s, s'_{i-1}) - f(s, s'_i)\right) \leqq 1$$

the right member of the last equation cannot be greater than M. That is

$$V_2 u(s, t) \leqq M. \tag{5}$$

It can be proved by the method Riesz uses in showing that a linear functional is defined uniquely for the limit of a sequence such as (4), that the function $u(s, t)$ just defined is the limit of $u(s+\epsilon, t)$ when $\epsilon = 0$. Thus $u(s, t)$ is regular.

It will now be proved that

$$U(f, g) = \int_T f(s)\, g(t)\, d_2 u(s, t), \tag{6}$$

where $f(s)$ and $g(t)$ are arbitrary continuous functions. The function $f'(s)$ will be defined by the equation

$$f'(s) = \sum_i f(s'_i)\left(f(s, s'_{i+1}) - f(s, s'_i)\right),$$

and $g'(t)$ in the analogous way. Then the equation

$$U(f', g') = \sum_{i,j} f(s'_i)\, g(t'_j)\, \Delta_{ij} u(s', t') \tag{7}$$

will be satisfied. It follows from definition that $f'(s) = f(s'_i)$ in the region $s'_i < s \leqq s'_{i+1}$, and similarly for g' and g. Since U is linear in each argument and f and g are continuous, $U(f'_i, g')$ approaches $U(f, g)$ when the length of the greatest of the intervals approaches zero, and the right member of equation (7) approaches the Stieltjes integral in equation (6). Inequalities (2) and (5) imply that $V_2 u$ is equal to M. This completes the proof that when $U(f, g)$ is bilinear there is a unique function $u(s, t)$ which is regular and satisfies equation (6), and the variation of u is the least upper bound of $|U(f, g)|/mfmg$.

If this property is assumed for functionals linear in each of $n-1$ arguments, the proof just outlined can be modified to make it prove that functionals linear in n arguments have the same property. Thus the theorem holds for n-linear functionals.

This can be used to extend a theorem of Fréchet's[3] about functionals of the second order to those of the nth order. A functional of the nth order is defined as one that is continuous and satisfies the equation

$$U(f_1+f_2+\ldots+f_{n+1})-\sum_{r=1}^{n+1}U(f_1+\ldots+f_{r-1}+f_{r+1}+\ldots+f_{n+1})+$$
$$\sum_{r=1}^{n}\sum_{s=r+1}^{n+1}U(f_1+\ldots+f_{r-1}+f_{r+1}+\ldots+f_{s-1}+f_{s+1}+\ldots+f_{n+1})$$
$$\ldots+(-1)^{n+1}U(0)=0$$

identically. If $U(f)$ is also homogeneous, the functional defined by the equation

$$W(f_1,f_2,\ldots,f_n)=U(f_1+\ldots+f_n)-\sum_{r=1}^{n}U(f_1+\ldots+$$
$$f_{r-1}+f_{r+1}+\ldots+f_n)+\ldots+(-1)^nU(0)$$

is easily proved to be linear in each of its arguments. If $f_1=f_2=\ldots=f$, it follows from definition and the condition of homogeneity, that

$$W(f,f,\ldots,f)=K_nU(f),$$

where

$$K_n=\sum_{r=0}^{n-1}(-1)^r\frac{n(n-1)\ldots(n-r+1)}{r!}(n-r)^n$$

But this expression can be proved equal to $n!$ which cannot vanish for any positive value of n, and since the n-linear functional W can be expressed as a multiple Stieltjes integral the homogeneous functional $U(f)$ of order n can be put in the same form.

[1] Riesz, *Ann. Sci. Ec. norm., Paris*, (Ser. 3), **28**, 1911, (36–43).
[2] *Ibid.*, **31**, 1914, (9–14).
[3] Fréchet, *New York, Trans. Amer. Math. Soc.*, **16**, 1915, (215–234).

THE CRYSTAL STRUCTURE OF CHALCOPYRITE DETERMINED BY X RAYS

By Charles L. Burdick and James H. Ellis

CHEMICAL LABORATORIES, THROOP COLLEGE OF TECHNOLOGY

Communicated by A. A. Noyes, September 29, 1917

Introduction.—This investigation of the atomic structure of crystals of chalcopyrite ($CuFeS_2$) was undertaken, as no study of a complex sulfide by the method of X-rays had previously been carried out. Moreover, comparatively few crystals of the tetragonal system, in which chalcopyrite crystallizes, have been examined; the only ones being certain oxides of the formula MO_2 studied by Vegard[1] and by Williams.[2] Yet the determination of the structure of crystals belong-

ing to other than the isometric system is likely to throw most light on the fundamental factors involved, such as the interatomic forces and the size and shape of the atoms themselves.

This research was carried out at the suggestion of Prof. A. A. Noyes with the aid of a grant made to him by the Carnegie Institution of Washington, for which we wish to express our indebtedness.

Description of the Apparatus and Procedure.—The method employed was in principle the same as that described by W. H. and W. L. Bragg,[3] but the apparatus was modified in several respects.

The electrical equipment for the X-ray excitation consisted of an interrupterless transformer of 20-kilowatts capacity. The X-ray tube was provided with a target of specially purified palladium, 2 mm. thick, silver-soldered onto a copper block, the face of which was set nearly perpendicular to the direction of the cathode rays. The electroscope used was one of the double-tilted form devised by Bumstead.

The crystal used in this investigation was selected from a number kindly loaned us by Prof. C. H. Warren of the Massachusetts Institute of Technology. The crystal was of the sphenoidal type,[4] approximately 8 mm. on the edges.

Summary of the Observations.—The following table summarizes the observations. The significance of the 'calculated' angles and intensities will be described below.

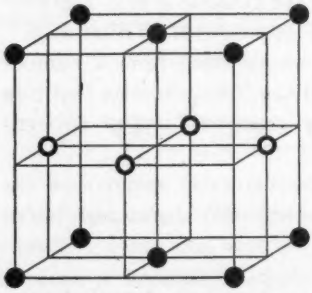
TABLE I

SUMMARY OF THE OBSERVED AND CALCULATED RESULTS

CRYSTAL PLANE	ORDER OF REFLECTION	ANGLE OF REFLECTION		ELECTROSCOPE DEFLECTION	RATIO OF INTENSITIES	
		Observed	Calculated		Observed	Calculated
(100)	I	6°25′	6°25′	5.5	100.0	100
	II	12°55′	12°55′	9.0	164.0	217
(001)	I	6°32′	6°32′	13.0	100.0	100
	II	13° 9′	13° 8′	33.5	256.0	222
(111)	I	5°32′	5°35′	49.0	100.0	100
	II	11° 7′	11°14′	6.5	13.0	4
	III	16°33′	16°58′	4.0	8.0	7
(11$\bar{1}$)	I	5°41′	5°35′	69.0	100.0	100
	II	11°26′	11°14′	10.0	14.0	4
(110)	I	9° 7′	9° 6′	190.0	100.0	100
	II	18°28′	18°27′	45.0	24.0	20
	III	27°52′	28°19′	11.0	6.0	7
(1$\bar{1}$0)	I	9° 4′	9° 6′	61.0	100.0	100
	II	18°23′	18°27′	29.0	47.0	20
(101)	I	9° 7′	9°10′	200.0	100.0	100
	II	18°29′	18°35′	50.0	25.0	20
	III	28°25′	28°32′	9.0	4.5	7

Interpretation of the Results.—In determining the atomic structure of chalcopyrite, we will first attempt to fix the type of lattice on which it is constructed. For approximate purposes its crystals may be regarded as isometric, since the axial ratio $a:c$ (1: 0.985) is nearly unity.

The interplanar distance d (that is, the distance between like planes of atoms) is by the fundamental law of reflection ($n\lambda = 2d \sin \theta$) inversely proportional to the sine of the angle of reflection (θ). For the three fundamental planes (100) (101) (111) the angles were 6°25′, 9°7′, and 5°32′. The reciprocals of the sines of these angles stand to one another in the ratio 1 : 0.706 : 1.152. The ratios of the interplanar distances in the three possible types of isometric lattice, known as the cubic, the face-centered, and the cube-centered[5] are respectively 1 : 0.707: 0.573; 1 : 0.707 : 1.146; and 1 : 1,414: 0.573. The lattice involved in the chalcopyrite crystal is therefore obviously the face-centered type; and we may infer that the heavy atoms, which predominate in determining the reflection, must be the points in this lattice.



COPPER AND IRON ATOMS IN CHALCOPYRITE LATTICE.

Since there is only a small difference in the atomic weights of iron and copper, their reflecting powers, like those of potassium and chlorine in potassium chloride, will not be greatly different. The atoms of iron and copper will therefore be practically indistinguishable in their effect on the X-rays. The observations then show that the basic lattice formed by the iron atoms must intersect that formed by the copper atoms in such a way as to form together a single face-centered lattice.

A study with the aid of a model of possible arrangements by which two different kinds of atoms, present in equal numbers, could together form a single face-centered lattice shows that there is only one such arrangement; namely, that shown in the accompanying figure, in which the iron atoms and copper atoms are represented by the solid and annular circles. It is evident, moreover, from the symmetries of the atomic arrangements that the vertical axis in the figure corresponds to the tetragonal or c axis in the crystal.

These conclusions are corroborated by the fact that the observed angles of reflections agree closely with the 'calculated' angles (given in the table), which were computed geometrically from the assumed loca-

tion of the atoms in the lattice, taking as a basis the observed angle for the (100) plane.

We may now proceed to determine the location of the sulfur atoms, which can be done with the aid of the intensity measurements. It is known that, when the reflection takes place from only one kind of plane, the reflections of the first, second, and third orders have intensities which stand to one another approximately in the ratios 100 : 20: 7, commonly called the normal intensity-ratios. It is evident, if there be secondary planes of lighter atoms intermediate between the primary planes of heavier atoms giving rise to a given reflection, that the reflection will be diminished or increased in intensity in correspondence with the difference in phase of the two trains of emerging rays waves. In case the secondary planes of atoms lie midway between the primary planes, there would be a phase-difference of half a wave-length for the first order, one wave-length for the second order, and $1\frac{1}{2}$ wave-lengths for the third order; hence there would result a weakening of the intensity of the first-order and third-order reflections (equal for the two orders) and a strengthening of that of the second order. In case the secondary planes are displaced one-fourth of the distance between the primary planes, the phase difference will be $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ of a wave-length for the first, second, and third orders, and hence there will be an increase in the intensities of the first order and third orders (equal in the two cases) and a relatively large decrease in the second order.

Let us consider now the relative intensities of the different orders for the different planes. We see that for the (100) plane the ratio of the observed intensities of the first and second orders has the value 100 : 164 in place of the normal one 100 : 20. This large relative weakening of the first order and strengthening of the second order shows that planes of sulfur atoms are located not far from midway between the (100) planes of iron and copper atoms. The reflections from the (001) plane show a similar reversal of the normal ratio; and lead to the corresponding conclusion that planes of sulfur atoms are located also not far from midway between the (001) planes of iron and copper atoms. The form and location of the lattice of sulfur atoms becomes thereby fixed, if we consider the additional fact that the number of sulfur atoms is equal to the number of iron and copper atoms. In the figure given above a sulfur atom would be located at the center of each alternate cube (that is, in four of the eight cubes there represented).

This conclusion is confirmed by the observations of the reflections from the other planes. The geometrical relations show that this location of the sulfur atoms involves that they lie in the (110) and (101)

planes of the iron and copper atoms; and that there is a plane of sulfur atoms in the (111) plane displaced one-fourth of the interplanar distance between the composite planes of iron and copper atoms. Correspondingly, the (110) and (101) planes show the normal ratios of intensities of the reflections for the first, second, and third orders; and the (111) plane shows the required decrease of the second-order reflection, and the expected normal ratio of the first-order and third-order reflections.

The relative intensities of the different orders can be calculated from the principle that the intensity of reflection from a plane of atoms is proportional to the square of the mass per unit-area.[6] Thus, the intensity of reflection from the (111) sulfur-atom plane would be to that from the (111) composite iron-copper atom plane as $(2 \times 32)^2$: $(56+63.6)$,[2] since there are two atoms of sulfur in an area equal to that in which there is one atom of iron and one atom of copper. The intensity of the resultant reflection will evidently be dependent both on the magnitudes of these two component reflections, and on the difference in phase in which they emerge. Algebraic expressions for the relative resultant intensities of the reflections of the different orders can be readily formulated. With the aid of these expressions the 'calculated' ratios of intensities given in the above table were obtained. It will be seen that there is a striking parallelism between the calculated and observed intensities.

Finally, we may further test the correctness of the deduced atomic structure by calculating the density of the substance and comparing it with the known density. Referring to the figure, it is seen that the space which it represents has associated with it two iron atoms, two copper atoms, and four sulfur atoms, or 2 of the atom-groups $CuFeS_2$. Since the mass of the hydrogen atom is 1.64×10^{-24} grams, that of these two $CuFeS_2$ groups is $2 \times \frac{183.6}{1.008}$ times as great, or 5.972×10^{-22} grams. The volume of the space in question is, however, equal to $8\,d^3 \times 0.985$, where d represents the distance between the (100) planes of copper-iron atoms (that between the (001) planes being $0.985\,d$). This distance d may be obtained from the law of reflection $\lambda = 2\,d \sin\theta$ by substituting for θ the observed angle of reflection (6°25′) for the (100) plane, and by substituting for λ its value 0.584×10^{-8} cm. as determined by W. L. Bragg[7] for a palladium target. The value of d is thus found to be 2.614×10^{-8} cm., and that of the volume in question 1.407×10^{-23} cm. The calculated density is therefore 5.972/1.407, or 4.24. The density of the mineral chalcopyrite, according to the

best determinations, lies between 4.1 and 4.3. That of the specimen used in this investigation is 4.19.

[1] Vegard, *Phil. Mag., London*, Nov., 1916 and Jan., 1917.
[2] Williams, *London, Proc. R. Soc.*, 93, 1917, (418).
[3] Braggs' *X-Rays and Crystal Structure*, 1915, (22).
[4] Dana's *Mineralogy*, Fig. 10, 1892, (81).
[5] Tutton's *Crystallography*, 1911, (501); or Braggs' *X-Rays and Crystal Structure*, 1915, (91).
[6] Bragg, *op. cit.*, pp. 120–127.
[7] Kaye, *X-Rays*, 1917, (226).

THE ISOSTATIC SUBSIDENCE OF VOLCANIC ISLANDS

By W. M. Davis

DEPARTMENT OF GEOLOGY AND GEOGRAPHY, HARVARD UNIVERSITY

Communicated October 9, 1917

Objection has sometimes been made to Darwin's theory of upgrowing coral reefs on subsiding foundations on the ground that the subsidence of the reef foundations should lower the ocean surface and lay bare a belt of recent marine deposits of smooth surface and simple shore line around the continental masses, while as a matter of fact the shore lines of continents are usually more or less embayed, as if the sea had recently advanced upon the unevenly eroded surface of the continental margins. The objection will, however, be found to have small weight when it is seen to rest upon the implied postulate that the embayments of continental coasts have as a rule been produced by a universal rise of the ocean surface, everywhere of the same amount and date, whereas their embayments testify to no such simple origin; and to proceed upon the unwarranted assumption that the subsidence of reef foundations requires the subsidence of broad areas of the ocean floors, whereas the local subsidence of the foundations themselves is all that is necessary. The postulate will first be examined and refuted; the assumption will then be considered and its alternative will be preferred.

The embayment of continental coasts is certainly of widespread occurrence, but when the embayments are closely examined they are found to be of dates and dimensions so diverse that they cannot be explained by a universal rise of the ocean. In the first place, the most pronounced embayments are the fiord coasts of quaternary glaciation; there is good reason to believe that the great troughs of such coasts were scoured out deep below sea level by huge glaciers, and that, far from the ocean having recently risen to submerge the troughs, the coasts have in several fiord regions risen from the ocean, as their elevated shore lines testify; these

coasts are embayed in spite of recent emergence, and cannot serve as witnesses to a recent rise of ocean level.

Furthermore, the supposition that coastal embayments result from a rise of the ocean around still-standing continental borders carries with it the implication that the coastal valleys, now submerged and embayed, were everywhere eroded during the same period of previous emergence, and hence that they should all be eroded to the same depth, and that their width should be proportionate to the weakness of the rocks in which they were carved during the emergence period. But as a matter of fact, the coastal embayments of the continents occupy valleys of depths and widths so various that they cannot be accounted for by these simple conditions: the diversity of the embayments demands many local and diverse movements of continental borders, even though a universal rise of ocean level has recently taken place, such as that which the final climatic changes at the close of the Glacial period should produce. The chief effect of such a rise of ocean level would only be to bring about a slight preponderance of embayed coasts along continental borders which had suffered local uplifts and depressions in about equal proportions.

Diverse movements of continental borders are demanded not only by the diversity of their embayments, but also by the not rare occurrence of emerged coastal plains of marine sediments. Thus the western side of the Adriatic is bordered by an emerged and more or less dissected coastal plain, while the eastern side is bordered by a partly submerged and elaborately embayed mountain border. Greece and the Bosphorus bear marks of recent submergence, but the coast of Palestine appears to bear marks of recent emergence. The small embayments of the Texas coastal plain cannot be explained as resulting from the same amount of submergence which produced the deep rias in the montainous coast of northwestern Spain; nor can the unlike valleys of those two coasts have been eroded during the same antecedent period of emergence. The elevated shore lines of the California coast have no equivalent on the volcanic islands of the Pacific. The recent submergence of the northeastern coast of New Caledonia following a long-enduring emergence, as indicated by the deep embayments of its strongly clift shores, demands a different succession of changes of level from those recorded on the coast of Peru or of Virginia.

The postulate that the embayed coasts of the world demand a universal rise of the ocean for their explanation is largely based upon the indoor study of coastal charts, and not upon the field study of the coasts themselves; and inasmuch as most of the coastal charts of the world have been made by hydrographers, who concerned themselves little

with the physiographic history of the coastal slope, such charts cannot be relied upon to present all the information that is pertinent to the problem under discussion. Thus if the coast of Maine were studied only from charts, its many embayments alternating with ragged promontories and outstanding islands, together with its small headland cliffs and its small bayhead deltas, might well be interpreted as indicating the recent submergence of a hilly region that had previously been eroded to subdued forms: but a study of that coast on the ground soon discovers that it bears a discontinuous cover of marine clays up to 100 or 200 feet above sea level; and one must therefore conclude that the embayments of today remain after the partial emergence of a previously more submerged region.

It occasionally happens, however, that a hydrographic chart presents evidence which suffices to correct a fase inference regarding recent submergence: thus one of the Solomon islands, which has a well-embayed shore line, is fronted for many miles by an elevated off-shore barrier reef, even crested and 80 or 100 feet in height: hence although the coastal embayments show that submergence has taken place, the reef shows that the last change of level was an emergence, less in amount than the previous submergence.

In view of these examples it is evident that the chart of an embayed coast, outside of the coral zone and not yet examined geologically, cannot be trusted as indicating that the last change of ocean level was upward. Therefore the inference that there has been a general rise of ocean level, independent of oscillations during the Glacial period, is not at present well supported; and until it is well supported, a certain amount of broad ocean-floor subsidence in the coral seas, as inferred by Darwin from his theory of coral reefs, may be regarded as not impossible, particularly as its effect in lowering the ocean may have been largely counteracted by equally broad ocean-floor uplifts in other regions.

Nevertheless, it is somewhat arbitrary to assume that the broad areas of ocean-floor subsidence should occur chiefly in the warmer zone, where coral reefs could be formed, and that the areas of compensating uplift should be chiefly in the cooler oceans, or in those parts of the warmer oceans where no islands rise. Hence the supposition of local subsidence of volcanic islands, which usually serve as reef foundations, deserves consideration as an alternative to broad ocean-floor subsidence. This supposition would be objected to by scientists of the older school, who believed that volcanoes occupy areas of elevation. Thus Guppy said: "In establishing the fact of the presence of active volcanoes in regions of barrier reefs and atolls, I shall be removing one of the principal stand-

points of the theory of subsidence;" he added that this "places the supporters of the theory of subsidence in a dilemna."[1] At about the same time Hickson wrote: "I am persuaded that . . . . the presence of such an atoll as Passiac [in North Celebes] so close to a region of quite recent and considerable volcanic activity is difficult to account for under this [subsidence] theory."[2] Murray went so far as to object to Darwin's theory of subsidence on the ground that even extinct volcanoes are not likely to subside. He wrote: "Generally speaking, all the volcanic regions which we know have in the main been areas of elevation, and we would expect the same to hold good in those vast and permanent hollows of the earth which are occupied by the waters of the ocean. . . . . Areas of local depression are to be looked for in the ocean basins on either side of and between groups of volcanic islands and atolls, and not on the very site of these islands."[3]

Much may be said against this obsolescent view, and in favor of the opposite view that volcanic action and subsidence may sometimes be closely associated. Strong testimony to this effect is offered by the repeated occurrence of volcanic eruptions in areas of subsidence during the geological evolution of Great Britain, as worked out by an exceptionally competent geologist whose conclusion is: "The study of the records of volcanic action in Britain proves beyond dispute that the volcanoes of past time have been active in areas of the earth's surface that were sinking and not rising. . . . . I do not wish to maintain that the downward movement was necessarily a consequence of volcanic ejections . . . . . but I have sometimes asked myself whether it was not possibly increased as a sequel to vigorous action."[4] An interesting piece of evidence concerning the subsidence of the great volcanic island of Hawaii during the later stages of its eruptive growth is furnished by Branner, who gives good reason for thinking that the deep canyons in the northeastern sector of the island have been eroded in part of an older, deeply dissected, and now partly submerged volcanic mass, the remainder of which has been overwhelmed and buried under the much more recent lava floods that cover the major area of the island.[5]

The newest discussion of this problem is by Molengraaf, who calls attention to the results of recent gravity determinations, from which it appears that the volcanic islands of the Pacific "as far as they have been studied are not isostatically compensated, and, without exception, show a larger or smaller positive anomaly of gravity. . . . . These volcanic islands, rising . . . . . as cones or groups of cones of considerable bulk, cannot always remain in existence; under the influence of gravity they will without exception yield and sink down slowly. . . . .

The yielding and slow sinking of the volcanic islands under the influence of gravity must be regarded as the cause of the downward movement of large amount and long duration which must be assumed in order to explain the formation of barrier-reefs and atolls in true oceanic regions."[6]

This appears to me an important suggestion, and one that is likely to remove the objections to Darwin's theory of coral reefs in so far as they are directed against a great subsidence of broad ocean-floor areas: for although Darwin himself was led by his coral-reef theory to infer the subsidence of such areas, it is clear from the original exposition of his theory that local subsidence of reef foundations will serve all its needs. It may be added that the accumulation of the great limestone masses of atolls upon slowly sinking volcanic foundations must aid and prolong their sinking; also that no comparable sinking of volcanic cones upon continents need be expected, not only because of the differences supposed to exist between the earth's crust in continental and oceanic areas, to which Molengraaf calls attention, but also because continental volcanoes suffer erosion, whereby their waste is carried away and widely distributed, while oceanic volcanoes, even if they rise for a time above sea level and suffer erosion, retain the waste from their summits on their flanks.

But it is particularly the relation of Molengraaf's hypothesis to changes of ocean level that I desire to emphasize. Let it be imagined that the ocean floor suffers no deformation apart from that associated with volcanic action, and that a thousand great volcanic cones are built up from it in the coral zone, one after the other and at such intervals of time that their formation stretches through the Tertiary and Posttertiary periods. The building of the first cone would cause a slight rise of ocean level. As the cone slowly subsided the ocean surface would return to its normal stand, were it not that the subsiding cone is reconstituted in an atoll as fast as it subsides,[7] and that other cones are built up as the first one sinks. Later formed cones would prolong the changes thus initiated, and the slow rise of the ocean would continue, particularly if the isostatic sinking of some of the cones were incompletely accomplished. Even if some of the cones sank so fast that reef-making corals could not reconstitute them, the net result of this process, after many cones had been built up and sunk again, would be, not a lowering of the ocean surface such as according to the usual interpretation of the theory of subsidence has been supposed to accompany the upgrowth of coral reefs, but a slow, small, and long continued rise of the ocean surface.

The rise of the ocean surface thus caused would be much less in total amount than the sinking suffered by any one of the volcanic reef founda-

tions, and hence less than the inferred great thickness of many atoll masses, but it would presumably be sufficient to cause a moderate preponderance of submergence on continental coasts which themselves suffer many diverse movements of upheaval and depression. It is not, however, to be supposed that general warpings and deformations of the ocean floor, upward and downward, should be left out of consideration; such movements have surely taken place to a less or greater degree, particularly in the western Pacific, where coral reefs border continental islands. The integrated effect of all these causes of change in the level of the ocean surface cannot now be determined, because so little is known regarding the various factors of the problem: but nothing in the little that is known and in the much more that may be fairly inferred should be regarded as discountenancing the theory of upgrowing reefs on subsiding foundations, essentially as Darwin supposed. His primary theory of coral reefs holds good, although his supplementary theory of broad ocean-floor subsidence needs modification.

[1] Guppy, H. B., *Scot. Geogr. Mag.*, **14**, 1888, (121–137); see p. 135, 136.
[2] Hickson, S. J., *A naturalist in Celebes*, London, 1889; see p. 42.
[3] Murray, J., *Proc. Roy. Soc. Edinb.*, **10**, 1880, (505–518); see p. 516.
[4] Geikie, Sir A., *The ancient volcanoes of Great Britain*, London, 1897; see vol. **2**, p. 470.
[5] Branner, J. C., *Amer. J. Sci.*, **16**, 1903, (301–316); see p. 301–303.
[6] Molengraaf, G. A. F., *Proc. k. Akad. Wet. Amsterdam*, **19**, 1916, (610–627); see p. 619–620.
[7] This statement depends on the fact, certified by chemists, that the withdrawal of limestone from solutions in water diminishes the water volume by only a small portion of the volume of the withdrawn limestone.

ON THE DEFORMATION OF AN N-CELL

By Oswald Veblen

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated by E. H. Moore, October 8, 1917

I propose to prove that *any* $(1-1)$ *continuous transformation of an* n*-cell and its boundary into themselves, which leaves all points of the boundary invariant, is a deformation.*

For the purposes of this proof the n-cell may be taken to be the interior of an n-dimensional cube. A deformation is a $(1-1)$ continuous transformation F_1 which is a member (corresponding to $x = 1$) of a one-parameter continuous family of $(1-1)$ continuous transformations F_x $(0 \leqq x \leqq 1)$ such that F_0 is the identity. It is understood that each F_x is a transformation of the n-dimensional cube into a set of points of an n-dimensional Euclidean space in which the n-dimensional cube is situated.

The theorem is easy in the one-dimensional case. It has been proved in the two-dimensional case by H. Tietze (*Palermo, Rend., Circ. Mat.*, **38,** 1914, p. 247) and more simply, by H. L. Smith (in an article soon to appear in the Annals of Mathematics). No proof has been published so far as I am aware, for the higher cases.

The proofs by Tietze and Smith establish a stronger theorem than that stated above, for they show the existence of a family of transformations F_x each of which carries the square into itself and leaves all points of the boundary invariant. This restriction on the transformations F_x, that each of them shall carry the square into itself, is not needed in some of the important applications of the theorem; and without this restriction the theorem can be proved very easily.

The proof below is stated for the two-dimensional, but applies without change to the n-dimensional, case.

Let S_1 be a square, $ABCD$, whose sides are of length unity, it being understood that a square, unlike a cell, includes its boundary. Let T_2 be a translation parallel to the side AB which carries the side AD into the side BC, T_3 a translation parallel to the side BC which carries the side AB into the side DC, and T_4 the resultant of T_2 and T_3. Let S_2, S_3, S_4 be the squares into which S_1 is carried by T_2, T_3, T_4 respectively. Thus S_1, S_2, S_3, and S_4 together constitute a square whose sides are of length 2.

Let F_1 be a $(1-1)$ continuous transformation of S_1 into itself which leaves all points of the boundary of S_1 invariant. The transformation T_2F_1 (the resultant of F_1 followed by T_2) carries S_1 into S_2. I shall first show that T_2F_1 is a deformation and it then follows easily that F_1 is also a deformation.

The rectangle composed of S_1 and S_2 can be carried into the rectangle composed of S_3 and S_4 by a transformation Λ which for points of S_1, is the same as T_3 and for points of S_2, is the same as $T_4F_1^{-1}T_2^{-1}$. Since T_3 and $T_4F_1^{-1}T_2^{-1}$ have the same effect on the common points of the boundaries of S_1 and S_2, the transformation Λ is uniquely defined, $(1-1)$, and continuous.

The transformation $\Lambda.T_2F_1.\Lambda^{-1}$, as applied to S_3 is the same as $T_4F_1^{-1}T_2^{-1}T_2F_1T_3^{-1}$, which is the translation $T_4T_3^{-1} = T_2$, carrying S_3 into S_4; denote this translation T_2 by T_1. Let T_x $(0 \leqq x \leqq 1)$ denote the translation carrying S_3 a distance x in the direction of translation of T_1.

The existence of the family of translations T_x $(0 \leqq x \leqq 1)$ shows that T_1 is a deformation. But since $\Lambda.T_2F_1.\Lambda^{-1} = T_1$, $\Lambda^{-1}T_1\Lambda = T_2F_1$. Hence the existence of the set of transformations $\Lambda^{-1}T_x\Lambda$ $(0 \leqq x \leqq 1)$ shows that T_2F_1 is a deformation.

Any one of the transformations $\Lambda^{-1}T_x\Lambda$ effects a translation on the three sides, AB, CD, DA of S_1 and carries the side BC into the curve to which T_2F_1 carries the linear segment in which the square is met by a line parallel to AD at a distance x from AD. Since if T_x is the translation which has the same effect as $\Lambda^{-1}T_x\Lambda$ on the point A, the transformation $T_x^{-1}\Lambda^{-1}T_x\Lambda$ leaves all points of the three edges AB, CD, DA of S_1 invariant. Denote $T_x^{-1}\Lambda^{-1}T_x\Lambda$ by F_x. The set of transformations F_x $(0 \leqq x \leqq 1)$ is obviously a continuous one-parameter family of $(1-1)$ continuous transformations; F_0 is the identity; and F_1 the given transformation already denoted by F_1. Hence F_1 is a deformation.

The last paragraph can be replaced by the observation that since the product of two deformations is a deformation, F_1, which is the product of T_2^{-1} and T_2F_1, must be a deformation. It seems worth while, however, to indicate, as has been done, something of the nature of the family of transformations F_x which the process sets up.

A THEOREM ON SERIES OF ORTHOGONAL FUNCTIONS WITH AN APPLICATION TO STURM-LIOUVILLE SERIES

By George D. Birkhoff

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated by E. H. Moore, October 10, 1917

1. *The Theorem.*—An infinite set of continuous functions $u_1(x), u_2(x), \ldots$ is *closed* on the interval $0 \leqq x \leqq 1$ if there exists no continuous function $f(x)$ not identically zero for which $\int_0^1 f(x)\,u_n(x)\,dx$ vanishes for all n; the set is *normalized* if $\int_0^1 u_n^2(x)\,dx = 1$ for all n; it is *orthogonal* if $\int_0^1 u_m(x)\,u_n(x)\,dx = 0$ for $m \neq n$. Most of the series of mathematical physics are linear in closed normalized orthogonal sets of functions.

THEOREM. *If $u_1(x), u_2(x), \ldots$ form a closed normalized orthogonal set of functions, and if $\bar{u}_1(x), \bar{u}_2(x), \ldots$ form a second normalized orthogonal set such that*

$$\sum_{n=1}^{\infty} (u_n(x) - \bar{u}_n(x))\,u_n(y) \qquad (0 \leqq x, y \leqq 1)$$

converges to a function $H(x, y)$ less than 1 in numerical magnitude in such wise that the series multiplied through by an arbitrary continuous function $f(x)$ can be integrated term by term as to x and yields a uniformly convergent series, then the set $\bar{u}_1(x), \bar{u}_2(x), \ldots$ is closed also.

Proof. If the set $\bar{u}_1(x), \bar{u}_2(x), \ldots$ is not closed there exists an f not identically zero such that $\int_0^1 f(x)\,\bar{u}_n(x)\,dx$ vanishes for all n. In

this event, if we multiply through the equation of definition for H by $f(x)$ and integrate as we may by hypothesis, we find

$$\int_0^1 f(x)\, H(x, y)\, dx = \sum_{n=1}^{\infty} \int_0^1 f(x)\, u_n(x)\, dx \cdot u_n(y),$$

where the series on the right-hand side will converge uniformly by hypothesis and so represents a continuous function.

This series on the right has the value $f(y)$. In fact the difference

$$f(y) - \sum_{n=1}^{\infty} \int_0^1 f(x)\, u_n(x)\, dx \cdot u_n(y)$$

is a continuous function $\varphi(y)$ such that $\int_0^1 \varphi(y) u_n(y)\, dy$ vanishes for all n, precisely because the set $u_1(x), u_2(x), \ldots$ is normalized orthogonal. But the set $u_1(x), u_2(x), \ldots$ is closed by hypothesis; hence we infer that φ vanishes identically. Hence the right-hand member of the preceding equation has the value $f(y)$. That equation may now be written

$$\int_0^1 f(x)\, H(x, y)\, dx = f(y).$$

If the maximum numerical value of $f(x)$ for $0 \leqq x \leqq 1$ is $F \neq 0$, and if this value is taken on for $y = y_0$, we get

$$\left| \int_0^1 f(x)\, H(x, y_0)\, dx \right| = F.$$

But this is impossible since by hypothesis $|H| < 1$ and $|f| \leqq F$.

The set $\bar{u}_1(x), \bar{u}_2(x), \ldots$ is therefore closed also.

2. *An Application.* The Sturm-Liouville series, in a specialized but typical form, arises from the set of functions $u_1(x), u_2(x), \ldots$ which are the solutions of a linear differential equation of the second order

$$u'' + (\lambda - g(x))\, u = 0$$

satisfying boundary conditions $u(0) = u(1) = 0$, for the ordered set of parameter values $\lambda_1, \lambda_2, \ldots$ of λ respectively. We will assume that $g(x)$ and $dg(x)/dx$ are real and continuous.

By means of the methods of Sturm it is proved that $u_1(x), u_2(x), \ldots$ may be taken to form a normalized orthogonal set in which $u_n(x)$ will vanish precisely n times within the interval $0 \leqq x \leqq 1$. And the methods of Liouville give the asymptotic form of $u_n(x)$; it will be convenient for us to use the formula

$$u_n(x) = \sqrt{2}\left[\sin n\pi x + \frac{\varphi(x) \cos n\pi x}{n} + \frac{M_n(x, g(x))}{n^2}\right]$$

in which M_n is a continuous functional of x and $g(x)$, bounded as long as $g(x)$ and $dg(x)/dx$ are bounded uniformly for all n.

For the satisfactory investigation of the representation of an arbitrary function in Sturm-Liouville series it is necessary to know further that the set $u_1(x), u_2(x), \ldots$ thus obtained is closed. A very simple proof and the first is due to Stekloff.[1] His proof uses, however, the theory of functions of a complex variable. There is indeed a special case, namely the case when $g(x)$ vanishes identically and $u_n(x) = \sqrt{2} \sin n\pi x$, when a short proof is possible by an elementary method.

The theorem proved above makes possible an immediate demonstration that the set is closed, once the asymptotic formula above and the fact for the special case are granted. Let us replace $g(x)$ by $\sigma g(x)$ where σ is a real parameter varying from 0 to 1. If then $u_1(x), u_2(x), \ldots$ denote the members of the set for any particular σ and $\bar{u}_1(x), \bar{u}_2(x), \ldots$ denote the members for another value of σ, say $\bar{\sigma}$, it will suffice to show that the series for H will have the properties demanded in the theorem for $|\sigma - \bar{\sigma}| \leqq \delta > 0$. For in this event since the set is closed for $\sigma = 0$ (the special case), it will be closed for $\sigma \leqq \delta$ by an application of the theorem; by successive further applications of the theorem it is shown in the same way that the set is closed for $\sigma \leqq 2\delta, \sigma \leqq 3\delta, \ldots$, so that finally we infer that the theorem is true for $\sigma = 1$, i.e. for the given set.

In virtue of the explicit formula for $u_n(x)$ we have

$$u_n(x) - \bar{u}_n(x) = \sqrt{2}\left[\frac{(\sigma - \bar{\sigma})\,\varphi(x)\cos n\pi x}{n} + \frac{M_n(x, \sigma g(x)) - M_n(x, \bar{\sigma} g(x))}{n^2}\right],$$

$$u_n(y) = \sqrt{2}\left[\sin n\pi y + \frac{\varphi(y)\cos n\pi y}{n} + \frac{M(y, \sigma g(x))}{n^2}\right].$$

Multiplying these two expressions on the right together we obtain the typical term of the H series. This series breaks up at once into six (2×3) other series all of which converge uniformly to a small value for $|\sigma - \bar{\sigma}|$ small save the series obtained from the leading terms in both expressions, namely

$$2(\sigma - \bar{\sigma})\,\varphi(x) \sum_{n=1}^{\infty} \frac{\cos n\pi x \sin n\pi y}{n}.$$

If we omit the small factor $(\sigma - \bar{\sigma})\,\varphi(x)$, this series may be written

$$\sum_{n=1}^{\infty} \frac{\sin n\pi(x+y) - \sin n\pi(x-y)}{n}.$$

Now the sum of any number of terms of a series $\Sigma (\sin n\pi z)/n$ is known to remain uniformly bounded,[2] and the series is known to converge everywhere, uniformly save in the immediate vicinity of the values $z = 0, \pm 2\pi, \pm 4\pi$, etc. Hence the series displayed above converges uniformly for all x, y save in the immediate vicinity of $x = y$, where, however, the sum of any number of terms of the series is bounded. Hence all six types of series will converge to values small numerically for $|\sigma - \bar{\sigma}|$ small, and, when multiplied through by any continuous function may be integrated term by term as to x, yielding uniformly convergent series in y. Thus the H series will have the stated properties for $|\sigma - \bar{\sigma}| \leqq \delta > 0$.

3. *Generalization.*—The theorem suggests at once a theorem in General Analysis as defined by E. H. Moore.[3] If we employ a quasi-geometrical terminology this generalization may be stated as follows: *any set of orthogonal vectors in a functional space lying near enough to a complete set of orthogonal vectors in that space is itself complete.* Another still wider generalization suggests itself: *any set of vectors in a functional space lying near enough to a complete set of vectors admitting a reciprocal set is itself complete and admits a reciprocal set.*[4] This second generalization evidently plays the same part in relation to biorthogonal sets that the first does for orthogonal sets.

[1] See Kneser, *Die Integralgleichungen und ihre Anwendungen in der mathematischen Physik*, Braunschweig, 1911, pp. 84–95.

[2] See D. Jackson, *Rend. Circ. Mat. Palermo*, **32**, 1911, (257–262).

[3] See *Bull. Amer. Math. Soc., New York*, **18**, 1911–1912, (334–362).

[4] For a theorem of this type see an article of mine, *Paris, C.-R. Acad. Sci.*, **161**, 1917, (942–945).

LOW-TEMPERATURE FORMATION OF ALKALINE FELDSPARS IN LIMESTONE

By Reginald A. Daly

DEPARTMENT OF GEOLOGY AND GEOGRAPHY, HARVARD UNIVERSITY

Communicated by W. M. Davis, October 5, 1917

Recent monographs by A. Heim[1] and D. Trümpy,[2] dealing with certain rock formations in Switzerland, put new emphasis on an important problem in minerogenesis. At different horizons of the Jurassic limestones of the Churfirsten-Mattstock mountain group, Heim has found abundant crystals of albite which have evidently developed *in situ* and are not of clastic origin. The crystals are automorphic, with maximum lengths of 0.2 mm. and average lengths much less (see fig. 150 in Heim's memoir). Minute crystals of ankerite are associated. Both albite and ankerite are regarded by Heim as due to crystallization on the sea-floor, during

the deposition of the limestone mud, which he interprets as likewise a chemical precipitate. The calcareous matrix has, in fact, the uniform, exceedingly fine grain characteristic of chemically precipitated limestone, though it encloses fragments of fossil shells and skeletons.

Trümpy, in the western Rhätikon, found similar albites, new crystallizations, in Jurassic limestone. These crystals are generally less than 0.03 mm. long and seldom as much as 0.05 mm. long. They are associated with colorless to somewhat brownish foils of mica with the same order of dimensions. Trümpy also describes secondary albite and orthoclase crystals, 0.05 to 0.08 mm. long, in the Tertiary (Flysch) limestones of the same district. The feldspars enclose roundish granules of calcite identical in size and form with the normal constituents of the mud-like limestone. In many cases the feldspar is merely a 'Bindemittel' for such grains of calcite. Trümpy agrees with Heim that the feldspars grew, 'schwebend,' in the calcareous mud.

Both Heim and Trümpy found no difficulty in distinguishing these new feldspars from the quite different clastic feldspars occasionally seen in the same thin sections. They have good evidence that the new feldspars were not developed by dynamic or contact metamorphism of the limestones—a conclusion borne out by the extremely fine grain of the calcareous material.

Thirty years before, Kaufmann[3] described similar albites, "doubtless formed *in situ*", as occurring in Jurassic, Lower Cretaceous, Upper Cretaceous, and Eocene limestones of the "Emmen- und Schlierengegenden" of Switzerland. The maximum length of the albite crystals in any of the formations is 0.08 mm. Kaufmann noted that they abound specially in the purer limestones.

Independently Lory[4] had found non-clastic albite and orthoclase crystals in most of the non-metamorphosed Jurassic limestones of the French Alps. Such feldspars are described as particularly abundant in nodular and geodic concretions. Pyrite crystals and minute, doubly-terminated quartz crystals accompany the feldspars. All the crystals are explained as growths in limestone muds charged with organic (bituminous) matter. A year later Lory[5] noted the occurrence of albites and bipyramidal quartzes in Eocene limestone at Montricher, near Saint Jean-de-Maurienne. He attributed the formation of the feldspars to special conditions which were not stated, but he excluded metamorphism as an explanation. He found orthoclase crystals in the cast of an Ammonite.

Foullon[6] reported the fine-grained to dense Eocene limestone of the Aegean island of Rhodes to contain 1.1 to 1.2% of well crystallized sili-

cates, including very pure albite. The largest feldspars seen were from 2.5 to 3.0 mm. in length. He suggested that the albite crystals originated, as tiny individuals, at the contact of sea-water and the calcareous mud at the sea-bottom; and that these grew to present size in the loose sediment as deposition went on. Only on such a hypothesis, he held, could one explain the automorphic character of the crystals. Since "certainly considerable time was necessary for their formation before being completely enclosed in the mud," he regards these crystals as proving the slowness of the calcareous precipitation. Foullon notes the absence of appreciable lime in the feldspar molecule itself.

Cayeux[7] stated that similarly authigenic[7] orthoclase is to be found in all the named horizons of the Chalk of the Paris basin. The crystals are always minute, averaging 0.04 to 0.05 mm. With constant crystal forms, they occur singly, never in groups. It is observed that orthoclase and glauconite tend to exclude each other in the various limestones. Cayeux assumes a community of origin for both minerals, described as "secondary." After examining his material, Michel Lévy, Lacroix, and Termier all agree that the orthoclase must have formed *in situ*.

Grandjean[8] studied the chalk of Meudon, near Paris, concluding that its feldspar (probably microcline, rather than orthoclase) was formed on the floor of the Cretaceous sea, "contemporaneously with the deposition of the chalk." The feldspar "grew rapidly; it probably ceased to grow after it was once buried." He continued: "it is thoroughly [infiniment] probable that feldspar is forming on the present sea-bottom."

A review of these European studies leads to the following generalizations:

1. Authigenic alkaline feldspars appear in non-metamorphosed limestones of Triassic, Jurassic, Cretaceous, and Eocene age and specially characterize many horizons in the Alpine Jurassic and Paris-basin Cretaceous.

2. Albite or orthoclase may occur alone or they may be associated in the same layer of limestone. Pyrite, bituminous matter, and bipyramidal quartz are not uncommon companions of the feldspars. One Jurassic limestone exhibits authigenic mica.

3. These feldspars in the European limestones typically appear as single crystals, with good crystal form and separated by the calcareous matrix; their lengths range generally from 0.03 to 0.08 mm.

4. All the calcareous matrices have the grain and texture of chemical precipitates. However, embedded in those originally muddy materials are the strongly contrasted fragments of organic origin, shells and skeletons.

5. All ten of the authors who have reported on the rocks mentioned agree that the albite or orthoclase has crystallized *in situ* without any help from metamorphism. Foullon, Grandjean, Trümpy, and Heim definitely express the view that the feldspars developed, as *chemical precipitates on the sea-floor* before or immediately after the burial of the corresponding laminae of sediment by the younger laminae. According to Cayeux, Lory held the same belief for the orthoclase crystals of the Jurassic limestones. The consensus of opinion is, then, that the authigenic feldspars and quartz attained their full growth well before normal diagenesis was completed, that is, while the calcareous muds were still unconsolidated.

These European researches are important from several points of view. For students of sediments they indicate a group of facts and problems which have been almost wholly untouched in American laboratories. They seem to throw light on the origin of an important dolomite in the Rocky Mountains. Finally, they furnish proofs of the crystallization of albite and orthoclase at temperatures much lower than have generally been assumed as possible.

In order to complete the summary of the recorded cases, that of the Rocky Mountain dolomite may be noted. It appears to represent the only known American parallel to the peculiar European limestones. The dolomite, named the Waterton formation, was discovered in 1905 by the writer,[9] at Waterton lake just north of the boundary between Montana and Alberta. The formation, at least 60 meters thick, is the lowest exposed member of the great Rocky Mountain geosynclinal and is overlain by 4000 meters of other sediments referred to the Beltian (and probably Cambrian) series. Many hundreds of meters of still younger strata overlay those rocks before the composite geosynclinal was finally upturned, in the Tertiary era. The whole sedimentary cover on the Waterton dolomite was probably more than 6000 meters in thickness; yet the dolomite is lacking in signs of metamorphism, even that of the static kind. Igneous metamorphism can be excluded with equal definiteness.

The dolomite is chiefly composed of carbonate grains, which average about 0.02 mm. in diameter and seldom reach 0.05 mm. in diameter. These grains either interlock or occur as sharp rhombohedra, representing the crystal form so characteristic of true dolomite. Many laminae of the rock, up to 1 mm. or more in thickness, are charged with disseminated grains of glass-clear orthoclase, 0.01 to 0.05 mm. in diameter and without good crystal form. A few irregular grains of quartz, with similar dimensions, are usually associated. Other laminae are very rich in

clumps and lenticular masses of similar, minute crystals of orthoclase, which are interlocked and thus without crystal form (see fig. 7 of the writer's memoir). The clumps and lenses of feldspars regularly enclose many rhombohedra of dolomite. Occasionally in thin section one sees entire laminae, up to 1 or 2 mm. in thickness, composed of about equal parts of orthoclase and dolomite. The feldspar is clearly not of clastic origin.

Analysis of the rock gave the following result, which has been thoroughly checked, the alkalies being subjected to four closely accordant determinations by two different analysts.

SiO_2	30.46	K_2O	5.77
Al_2O_3	6.86	H_2O	1.42
Fe_2O_3	4.53	CO_2	22.55
FeO	1.89		
MgO	10.07		99.95
CaO	16.02	Sp. gr.	2.749
Na_2O	0.38		

The estimated composition of the rock (assuming the iron oxides to represent magnetite) is, by weight:

	per cent		per cent
Orthoclase	34.5	$MgCO_3$	21.2
Albite	3.1	$CaCO_3$	28.6
Quartz	6.0		
Magnetite	6.3		99.7

The albite is probably in solid solution with the orthoclase. So far as known by published analysis no other non-metamorphosed dolomite or limestone even approaches the Waterton dolomite in its abundance of the alkalies; feldspar makes up about 40% of its volume.

When reporting on this singular rock the writer was not aware of the relevant observations of Lory, Foullon, Cayeux, and others. In 1912 the dolomite remained frankly a puzzle. The theoretical conclusions of Heim, Trümpy, and Foullon are therefore of value in suggesting a partial explanation of the still unique American rock. The Alberta formation differs from the European analogues in having a dolomitic base and in failing to exhibit good crystal form for the feldspar. This anhedral character may be connected with the specially high proportion of orthoclase, the numerous little crystals interfering with one another as they grew in the calcareous mud. The writer has not the slightest doubt as to the chemical origin of the carbonate material nor as to the authigenic nature of the orthoclase; he is inclined to refer the crystallization of the feldspar to the early period when the sediment was not yet consolidated nor deeply buried.

If Foullon, Grandjean, Trümpy, and Heim are correct in regarding these albite and orthoclase crystals as chemically precipitated *at* the sea floor, the formation temperatures must have been well below 100°C. If, on the other hand, the crystals developed only after burial of their calcareous matrices, the formation temperatures were nevertheless comparatively low. The Paris-basin rocks were never buried more than a few hundred meters. Assuming a burial of 1000 meters and the present thermal gradient (3° per 100 meters of depth), the formation temperatures could hardly have reached 100°. Though the burial was deeper for other cases (Alberta and the Alps), formation temperatures of much more than 200° cannot well be assumed, even if the crystallization were delayed until the maximum sedimentary covers were completed.

In any case the Paris-basin limestones apparently prove that albite, orthoclase, and quartz are crystallized in calcareous muds at temperatures lower than 100° and probably lower than 70°. Since this paper was practically completed for publication, it was found that Doelter[10] had already assumed a formation temperature of 'perhaps 100°' for the potash feldspars of the Paris-basin Chalk.

The lowest recorded temperature at which orthoclase has been artificially prepared is 300°. That feat was accomplished by Chrustschoff,[11] who developed both orthoclase and quartz in an aqueous solution of dialyzed silica with alumina and caustic potash, kept for several months at the temperature mentioned. Friedel and Sarasin[12] using analogous solutions in bombs, produced albite, orthoclase, and quartz in much shorter periods, but they employed a temperature of 500°. These experiments suggest that time may be one of the important factors aiding the crystallization of alkaline feldspars at very low temperatures.

What the exact chemical conditions for these noteworthy precipitations in calcareous mud were, is a question now impossible to answer. Perhaps the abundance of lime in the muddy matrices affected the solubility of the alkalies in sea-water;[13] one should further consider the possible influence of decaying animal matter on the various solubilities involved. The astonishing concentration of potash feldspar in the Waterton dolomite accentuates the difficulty of the problem.

[1] Heim, A., *Beitr. geol. Karte Schweiz*, N. F., Lief **20**, 1916, (514, 543, 561, 567).
[2] Trümpy, D., *Ibid.*, Lief., **46**, 1916, (83, 108).
[3] Kaufmann, F. J., *Ibid.*, (*Ser.* 1), Lief., **24**, 1886, (583).
[4] Lory, C., *Paris, C.-R. Acad. Sci.*, **103**, 1886, (309).
[5] Lory, C., *Ibid.*, **105**, 1887, (99).
[6] Foullon, H. B., *Sitzber. Wiener Ak.*, **100**, Abt., 1, 1891, (162, 169).
[7] Cayeux, L., *Paris Mém, Soc. Géol.*, **4**, No. 2, 1897, (259, 279, 303, 432–434). F. Grandjean, *Paris, C.-R. Acad. Sci.*, **148**, 1908, (723), holds that these potash feldspars described

by Cayeux have the optical properties of microcline. Termier found authigenic albite crystals in the Flysch limestone of Briançonnais. Compare Rosenbusch, H., *Elemente der Gesteinslehre*, 3te Aufl., Stuttgart, 1910, (520). "Authigenic" means "formed *in situ*."

[8] Grandjean, F., *Paris, C.-R., Acad. Sci.* **148**, 1908, (723).

[9] Daly, R. A., *Geol. Surv. Can., Memoir* No. 38, 1912, (50).

[10] Doelter, C., *Handbuch der Mineralchemie*, Dresden, Bd. 2, 2te Hälfte, 1915, (556).

[11] Chrustschoff, K., *C.-R. Acad. Sci.*, **104**, 1887, (602).

[12] Friedel, C., and Sarasin, E., *Ibid.*, **92**, 1881, (1374).

[13] Cf. Zschimmer, E., in C. Doelter's *Handbuch der Mineralchemie*, **1**, 1912, (910).

THE INTERFEROMETRY OF SMALL ANGLES, ETC. METHODS BY DIRECT AND REVERSED SUPERPOSED SPECTRA

By Carl Barus

DEPARTMENT OF PHYSICS, BROWN UNIVERSITY*

Communicated September 21, 1917

1. Introductory.—It occurred to me that a number of the methods treated in my papers on direct and reversed spectrum interferometry might be used directly for the measurement of small angles and possibly of the distance of the source of light. Such a procedure would have an apparent advantage, at least theoretically, of not calling for the preliminary superposition of the images of distant objects, as the superposition is inherent in the method itself. But there are large constants involved which make the result very problematical, unless these constants can be removed by a compensator. It is very questionable, moreover, whether, appreciable interferences can occur, and another difficulty which hampers the method is the decrease in the size of objects as their distances increase. A progressive investigation with the object of ascertaining to what degree the experiment is feasible is nevertheless worth while. It will be convenient therefore first to develop the methods without reference to the ulterior conditions which limit the interferences and this method has been pursued.

2. Method with Prism.—Figure 1 is a sketch of one of the methods in which S is the distant source of light, from which rays d and d' strike the mirrors m and n, are thence reflected to the silvered sides of the right angled prism P. After leaving it the rays enter the spectro-telescope at T in parallel. If the proper angles are selected the prism P may be replaced by one of any angle or by a reflecting grating.

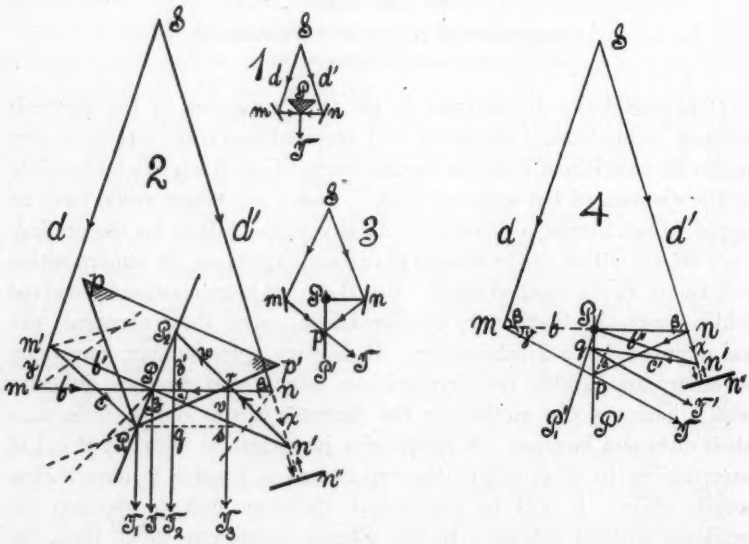
Suppose now the system mPn is securely attached to a rigid metallic beam or rail capable of rotating around a vertical axis at its center (P). This is indicated in figure 2 where the direction of rays and the normals

* Advance note from a Report to the Carnegie Institution of Washington, D. C.

of mirrors have been drawn and where the angle of rotation α has placed mPn into the position $m'Pn'$. The result is that a part y of the ray d is cut off on the left side and a part x added to the ray d' on the right side, so that the path difference which may be assumed to have been zero originally is now appreciably incremented, but not symmetrically for both sides.

It may be shown however that the rays $n'P_2T_2$ and $m'P_1T_1$ still enter the telescope in parallel and that therefore the conditions of interference have not been disturbed. This is the interesting feature of the method.

It will contribute to a more adaptable design of the apparatus for



general interferometry, if the ray Sn'' may also be reversed by reflection parallel to itself from a normal mirror n'', allowing a small lateral offset, similar on both sides for clearance of the mirrors. Again half silvers may be used at m and n for transmission and reflection, which method is probably best. These details will here be disregarded. If small angles are to be measured the direct ray method is enormously more sensitive.

3. Equations.—To derive the equations certain intercepts of the rays figure 2 in addition to x and y, may be defined. P_1P_2 is the trace of the vertical plane of symmetry of the right angled prism, if rotated at an angle α to the right. In this case the reflected ray $n'P_2q$ on the

right corresponds to the reflected ray $m'P_1$ on the left both terminating in the common wave front P_1qs before entering the telescope.

$$\begin{aligned} \text{Let } n'P_2 &= c' = b \sin\beta \,/\, \cos\alpha \sin(\beta - \alpha) \\ m'P_1 &= c = b \sin\beta \,/\, \cos\alpha \sin(\beta + \alpha) \\ P_2t &= z' = b \sin\alpha \sin\beta \,/\, \sin(\beta - \alpha) \\ tq &= z = b \sin\alpha \sin\beta \,/\, \sin(\beta + \alpha) \\ \text{and } nn' &= x = b \sin\alpha \,/\, \sin(\beta - \alpha) \\ mm' &= y = b \sin\alpha \,/\, \sin(\beta + \alpha) \end{aligned}$$

since the orginal angles at the ends of the base are β and the rotation α. The angles between incident and reflected rays are respectively $\beta - 2\alpha$ at n', $\beta + 2\alpha$ at m', $90° - 2\alpha$ at P_2 and $90° + 2\alpha$ at P_1. Finally b has changed to b' on the left and b'' on the right.

The rays however do not reach the plane of symmetry but are reflected by the faces of the right angled prism and this may be sketched in apart from size, in the rotated position (angle α) at P_1pp'. The path of the reflected rays from n' is now $n'rs$ and from m', $m'P_1$, before they meet in the common wave front P_1qs. Hence the intercepts

$$\begin{aligned} rP_2 &= w = (z + z') \,/\, \cos\alpha\,(\sin\alpha + \cos\alpha) \\ rs &= v = (z + z')\,(\cos\alpha - \sin\alpha) \,/\, (\sin\alpha + \cos\alpha) \end{aligned}$$

will enter in treating the path differences. On the left the rays have not been disturbed.

If we take the direct case first the orignal path difference SnP and SmP may be regarded zero or n and m in the same phase. On rotation therefore (angle α) the path difference is equivalent to the equation

$$n\lambda = c' - c - (w - v) + x + y$$

If the above equivalents are inserted, this equation though very complicated, may to the second order of small quantities, be reduced to a form which, since α and $90° - \beta$ are small angles for practical purposes may be abbreviated to (n, order of interference).

$$n\lambda = 2b\alpha - 2b\alpha^2 + 2b^2\alpha \,/\, d$$

The three terms correspond to the xy, wv and cc' effect.

In the case of reversed ray (fig. 2) we may consider the points m' and n' in the same phase. Hence the original path difference ($\alpha = 0$) is $x - y$. The path difference after rotation $c' - w + v - c$. The total change of path difference due to rotation is thus given by

$$n\lambda = c' - c - (w - v) - x + y$$

This differs from the preceding by the deduction of $2x$. The rays again terminate in the common wave front P_1qs to enter the telescope. Here the long rigorous equation reduces practically to

$$n\,\lambda = 2\,b\,\alpha^2 - 2\,b^2\,\alpha\,/\,d$$

The ww effect predominates, the cc' effect is intermediate and the xy effect very small if d is large, as already instanced. If $\alpha = 1°$, $b = 1$ meter, $d = 1$ kilom., $\lambda = 6 \times 10^{-5}$ cm., the three terms of the first equation give respectively 6×10^4, 10^3, and 60 fringes.

To the first order of small quantities the equations may be written, $n\,\lambda = 2\,b\,\alpha\,(\text{cosec}\,\beta + \cot\beta)$ and $n\,\lambda = b\,2\,\alpha\cot\beta$. They may also be obtained geometrically by letting fall a normal from n' to the near prism face, prolonging T_3 backwards and using the isosceles triangle obtained. In this case $x + 2\,(c' - w)\cos^2(45° - \alpha) - (z' - z) + y - c$ is the path difference and reduces to the above equation. So long as the nose of the prism is near the axis of rotation, the equation of first order quantities need not be modified.

A very essential correction however is still needed. In the practical apparatus the mirrors m and n rotate on a rigid radius, b, whereas in the diagram (if $\delta = \beta - \alpha$ and $\sigma = \beta + \alpha$), b elongates on the right and contracts on the left to $b'' = b \sin\beta / \sin\delta$, and $b' = b \sin\beta / \sin\sigma$. Hence the mirrors on the right and left are displaced normally by $(b'' - b)\cos\beta/2$ and $(b - b')\cos\beta/2$. The path difference introduced is thus $(b'' - b)(\cos\alpha + \cos\delta) + (b - b')(\cos\alpha + \cos\sigma)$ which to the first order of small quantities may be written $2\,b\,\alpha\,(1/\sin\beta - \sin\beta + \cot\beta)$. If this quantity is deducted from the right of the above equation for path difference and direct rays there remains simply $\eta\,\lambda = 2\,b\,\alpha\sin\beta$. This therefore is the equation to be used in interpreting the observations so that generally $2\,\Delta N \cos i / \Delta\alpha = 2\,b\sin\beta$ where $i = \beta/2$ for the micrometer at n.

In the case of reversed rays the conditions on the left remain the same as before; whereas on the right the mirror is set at an angle $\beta/2$ to the rail. Hence the normal displacement is $(b'' - b)\sin\beta/2$ and the angle of incidence $90° - (\beta/2 - \alpha)$. Thus the full path difference here to be deducted is $(b'' - b)(\cos\alpha - \cos\delta) + 2\,(b - b')(\cos\alpha + \cos\sigma)$ which to the first order of small quantities reduces to $2\,b\,\alpha\cot\beta$. Hence the equation for reversed rays is simply $2\,\Delta N / \Delta\alpha = 0$ and we have an interesting appearance of terms of the second order only, which I will here omit. In general glass paths may be compensated at pleasure.

The *observations* made with the present apparatus, though quite interesting, are beyond the purpose of the present note. The two

corresponding rays d and d' were obtained by cleaving the white beam from a collimator symmetrically, by a knife edged prism with silvered faces.

4. Interferences from Rough Surfaces.—The question now at issue is whether these interferences can be retained when the collimator is removed and the light comes directly from a ground glass surface or a Nernst filament. The spectrum fringes go at once when the slit is widened. Not so the achromatic sets. After obtaining this result with sunlight and ground glass, I replaced both by the light from a Nernst filament, under the impression that ground glass might, to a small degree, behave like plate glass. Having produced the achromatics as usual with the collimator, its objective was first removed and it was then seen that the two washed slit images were no longer superposed. Bringing the images together (by rotation of the mirror n around a vertical axis), a position was soon found in which the achromatic fringes appeared brilliantly in a white field, quite out of focus. The slit could now be widened or removed altogether, but the fringes persisted though with less brilliancy. It is thus possible to obtain these fringes directly from a Nernst filament or a narrow vertical strip of sunlight. They are so mobile with changes of Δ N and Δ α, that to find them it is necessary first to produce the spectrum fringes with collimator and spectro-telescope, then to find the achromatics on removing the spectroscope, next to remove the objective of the collimator and adjust for superposed images and finally to remove the slit. They practically cover the whole width of the washed slit image with streamers extending laterally into the glare some five times further. Slit images may even be slightly separated while each alone retains the achromatic fringes, a rather puzzling phenomenon. One may note that the slit images here are not reversed.

Experiments made as to the nature of these achromatic fringes showed that they are probably Fresnellian interferences. To prove this the objective of the collimator was removed and strong fringes obtained by passing the two vague images of the slit gradually over each other, horizontally. The fringes in this motion passed from horizontal maxima of size gradually to vertical hair lines as the images slid from contact of their nearer edges to contact of their further edges. The coarse fringes were even strongly present in the narrow dark gap between slit images prior to contact. The telescope was now focussed on the slit so that sharp linear images appeared. The fringe then vanished, but it appeared that the coarse fringes corresponded to coincident sharp slit images when observed out of focus, and the fine hair lines to sharp slit images far apart also seen out of focus. The whole

phenomenon thus depends on the distance apart of two sharp lines of light and the interferences are observable before or behind their focal plane.

5. Reversed Rays.—If the necessary excess of path on the right is compensated by a glass column on the left (usually 10 or 15 cm. long), the spectro-telescope on adjustment shows a field of concentric *half* ellipses, all terminating in a vertical axis. These interesting phenomena are unfortunately not available for measurement as the terminator is not sharp enough and as their motion is necessarily sluggish in view of the large excess of glass path on one side. The achromatic fringes are not producible. For practical work the glass path must be replaced by an air path obtained by aid of an *offset* consisting of two pairs of parallel plates at right angles to each other. These not merely compensate without changing the direction of rays, but on rotating the plates as a whole around a horizontal and a vertical axis, they additionally serve for producing ellipses of any size and of centering them. Achromatie fringes are now brilliantly producible. The variety of observations made will however have to be given elsewhere.

6. Plate Method.—In view of certain difficulties encountered in the use of reflecting prisms, in particular the loss of rays at the edge, the method of figure 3 enlarged in figure 4 was devised. In this the prism is replaced by a half silver plate PP'. Hence the rays issuing at S and reflected by the opaque mirrors at m and n, are thereafter respectively transmitted and reflected by the half silvered plate PP' and then reach the spectro-telescope at T together. When the path differences are sufficiently equal, elliptic interference fringes will be seen in the spectrum. When first found they are usually very fine straight lines; but they may be rectified by plate compensators in the beams d and d' or mp and np, though the operation is not easy. Leaving these details for further consideration elsewhere, the procedure for angular measurement may advantageously be treated here. For this purpose the half silver P and one opaque mirror, n for instance are mounted on a rigid bar with an axis at P. The other mirror m is to remain fixed. If the bar is now rotated over a small angle α (fig. 4), the mirror at n is displaced to n' and the ray Sn prolonged (intercept x) is now reflected from n' to q and thence along T' into the spectro-telescope, parallel to its original direction or to the other ray mp. Hence the interferences remain intact but many fringes pass during the transfer. The *persistance* of parallelism is easily seen to be the essential feature of the method.

To control the fringes either the mirror at n (or at m) may be displaced on a micrometer screw normal to itself, or the half silver plate at P may be displaced parallel to itself. If the angle of incidence at n is i and the normal displacement of n is e the path difference intro-

duced will be $2e \cos i$. Similarly if the normal displacement of the plate P is e' and the angle of incidence i', the path difference will be $2e' \cos i'$.

As in the preceding experiment the mirror at n may be a half silver, so that the ray, d, passes through it and may then be returned in its own path by a mirror at n'', on a fixed standard but provided with a micrometer. The displacement of this mirror parallel to itself over a distance e, introduces the path difference $2e$, while during this motion the rays np or $n'q$ are now stationary. Beams of light do not pass through each other and the interferences are kept at full intensity throughout. The glass path at n compensates the glass path PP' The air path excess $2nn''$ on the right must be specially compensated by an offset in d, as explained above.

7. *Equations.*—The rigorous equations for this case are cumbersome. If in figure 4, m and n are in the same phase and Pp is symmetrical, there will be no path difference at p. When Pn is rotated over an angle α into Pn', the path on the right becomes $nn' + n'q + qs$ while (ps, wave front) the path on the left remains mp as before. The path difference is thus the difference of these quantities to which however the increased glass path at PP' would have to be deducted. If the angle SnP is β and Pnp γ, the values of the branch paths may be found to be (since $nP = mP = b$) is $\beta - \alpha = \delta$ and $\gamma - \alpha = \tau$

$$mp = np = b \,/\, \cos \gamma$$
$$nn' = b \sin \alpha \,/\, \sin \delta$$
$$n'q = b \sin \beta \,/\, \sin \delta \sin \tau$$

Hence qs and the path difference are complicated expressions which need not be inserted here.

If α is small, so that differential expressions may be introduced, the rigorous equation (to an approximation of the second order in α) is finally equivalent to $n\lambda = b\alpha(1 + \cos(\beta - \gamma))/\sin\beta$. If $\beta = 90°$, $n\lambda = b\alpha + p\alpha\cos\gamma$, where p is the distance Pp. The same expressions may be obtained geometrically by prolonging $n'P$ and $T'q$ and treating the isosceles triangle produced.

For the case where the ray Sn prolonged returns on itself as from n'', in figure 4, n being a half silver plate, the quantity $nn' = 2x = 2b\alpha / \sin\delta$ must be deducted. Hence $n\lambda = b\alpha(1 - \cos(\beta - \gamma))/\sin\beta$. When $\gamma = 0$, this equation coincides with the case of the prism method apart from the factor 2.

It is finally necessary to apply the correction for the occurrence of a constant radius of rotation, whereby the mirror n is both rotated and displaced. If the distances $Pn = b$ and $Pn' = b''$, the normal displacement is $e(b'' - b)\cos(\beta - \gamma)/2$. The angle of incidence being

$(\beta + \gamma) / 2 - \alpha$, the corresponding partial path difference works out as $(b'' - b)$ $(\cos (\beta - \alpha) + \cos (\gamma - \alpha))$; or finally in full as $b\ \alpha$ $(\cos^2 \beta + \cos \beta \cos \gamma) / \sin \beta$. Subtracting this from the first equivalent of $n\ \lambda$ above, the equation to be used for non-reversed rays becomes $n \lambda = b \alpha (\sin \beta + \sin \gamma)$ when α is small. As the angle of incidence at the mirror at n is $(\beta + \gamma) / 2$, and if ΔN is the displacement of this micrometer, the practical equation is thus $2 \Delta N \cos (\beta + \gamma) / 2 = b (\sin \beta + \sin \gamma) \Delta \alpha$.

For reversed rays the normal displacement is $(b'' - b) \sin (\beta - \gamma) / 2$ and the path difference for small α therefore $b\ \alpha \cot \beta$ $(\cos (\gamma - \alpha) - \cos (\beta - \alpha))$. Subtracting this from the corresponding equivalent of $n \lambda$, the equation for reversed rays is thus $n \lambda = b \alpha (\sin \beta - \sin \gamma)$ or in the practical form as the incidence is now zero, $2 \Delta N = b (\sin \beta - \sin \gamma) \Delta \alpha$.

All equations contain the distance of the remote object at S, in $\sin \beta$, so that d occurs as a second order quantity.

Observations (also to be omitted here) were made in great variety with this apparatus, a knife edge prism at S cleaving the white beam from a collimator. To obtain strong full spectrum ellipses, the rays T and T' must not only be parallel but interpenetrate so far as possible. If p and q are a few millimeters apart all fringes vanish. For this adjustment the compensator is again convenient and for reversed rays it must be of the offset type described. One may notice that the slit images are here mirror images of each other. Nevertheless brilliant achromatics may be obtained even from rough surfaces, if narrow, by the succession of operations given above. Both these and the spectrum fringes are available for measurement. The achromatics are most serviceable if tranverse to the slit image as they then rise and fall with the play of the micrometer. To obtain them, the center of ellipses must be placed in the vertical through the telescopic field; but above or below it. In other words the spectrum fringes are to be horizontal. Even when (as in the preceding section in view of the half ellipses) horizontal spectrum fringes are precluded, the achromatic fringes will be found by adjusting as if the former were to appear. The best achromatics consist of but one or two fringes, sharply in black and white, with three or four much fainter fringes on either side. These occur frequently, but how to differentiate them systematically from the other groups of 10 or 20, more nearly uniform and therefore less serviceable fringes, I have yet to learn. All must be treated with caution, however, for they move through the field of the telescope so rapidly that if lost it is usually expeditious to seek for them again through the spectrum fringes.